



Published in final edited form as:

J Neurogenet. 2013 June ; 27(0): 1–4. doi:10.3109/01677063.2013.772176.

Presence of epilepsy-associated variants in large exome databases

Natalya S. Cherepanova, BA¹, Elizabeth Leslie, PhD¹, Polly J. Ferguson, MD¹, Michael J. Bamshad, MD², and Alexander G. Bassuk, MD, PhD^{1,*}

¹Department of Pediatrics, The University of Iowa, Iowa City, Iowa

²Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington

Abstract

Mutations in more than twenty genes have been found to cause idiopathic epilepsies, and screening for these variants could facilitate the clinical diagnosis of epilepsy. However, many of the studies that reported putative pathogenic variants for epilepsy tested a relatively small number of control samples making it more likely that a rare non-pathogenic variant could be mistaken as causal. To test the robustness of inferences based on small sample sizes, we investigated whether variants previously reported to cause epilepsy were present in the resequencing data from the large control populations of the 1000 Genomes Project and the NHLBI Exome Sequencing Project. A list of variants associated with epilepsy was compiled using a manual review of the literature for genes associated with epilepsy from a recent International League Against Epilepsy (ILAE) report and two comprehensive genetic studies. We checked for the presence of those variants in the 1000 Genomes Project database and the NHLBI Exome Variant Server (EVS). Of 208 epilepsy-associated variants that we identified from our literature review, only seven were found among 17 thousand chromosomes across 1000 Genomes and the EVS. Consistent with recent published reports, we also found many variants with predicted pathogenicity in epilepsy associated genes in the genomic databases. Our findings suggest that the 1000 Genomes and the EVS datasets may be a valuable resource of control data in research aimed at identifying genes for epilepsy specifically when the model predicts a highly penetrant allele. These databases also elucidate the array of genetic variation in putative epilepsy genes in the general population.

Introduction

Recent research into genetic causes of epilepsy has linked over twenty genes to idiopathic epilepsies, and the International League Against Epilepsy (ILAE) genetics commission recently published a report that discusses this emerging information in relation to the diagnosis and treatment of epilepsy (Ottman et al., 2010; Harkin et al., 2007; Mulley et al., 2005). Many of the studies referenced by the ILAE report evaluated potentially deleterious protein-coding variants in relatively small control groups. However, recent population genetic analyses have demonstrated that humans harbor an abundance of rare deleterious variation, with more than 80% of all coding variants having a frequency of one percent or less (Tennessen et al., 2012; Nelson et al., 2012). Moreover, Klassen et al. (2011) found an equal frequency of mutations in ion channels in individuals with sporadic idiopathic epilepsy and controls; accordingly, it seems possible that non-pathogenic variants present at a low to intermediate frequency (i.e., < 5%) in the general population could be missed by

*To whom correspondence should be addressed: University of Iowa Carver College of Medicine, 200 Hawkins Drive, Department of Pediatrics, Division of Neurology, 2502 JCP, Iowa City, IA, 52242, Phone: (319) 356-7726, Fax: 319-384-8818, alexander-bassuk@uiowa.edu.

screening a small number of control samples and thus be misinterpreted as causal for epilepsy. The recently made public 1000 Genomes Project database and the NHLBI Exome Variant Server (EVS) could potentially serve as a large source of control data and mitigate this limitation (The 1000 Genomes Project Consortium, 2010; Exome Variant Server).

We investigated whether variants that have been recommended by the ILAE as likely causal idiopathic epilepsy variants (Ottman et al., 2010; Harkin et al., 2007; Mulley et al., 2005) are present in either the 1000 Genomes Project database or the EVS. Out of 290 variants, only seven were present in the EVS, suggesting that the vast majority of mutations identified by the ILAE are likely causal.

Methods

We compiled a list of variants that have been reported to cause epilepsy from Ottman et al., 2010, Harkin et al., 2007., and Mulley et al., 2005, and checked for the presence of those variants in either the 1000 Genomes Project database or the EVS. The Exome Variant Server used the sequences of roughly 6,500 exomes and is a compilation of samples sequenced from a variety of studies of heart, lung, and blood disorders. We used the ESP6500 version of the Exome Variant Server. This release included samples from 2,203 African-Americans and 4,300 European-Americans, for a total of 13,006 chromosomes (Exome Variant Server). The 1000 Genomes Project aimed to identify variants that occur at a frequency of 1% or greater in the population studied. It sampled a wide range of populations and currently has the sequences of 2,200 individuals available (The 1000 Genomes Project Consortium, 2010).

Results

We compiled a list of 290 variants among 19 different genes associated with epilepsy (Table 1). Variants were typically identified in only a single individual, or were private to individuals with epilepsy in large multiplex families. Of those, 82 (28.3%) were indels, frameshifts, or splice site variants, and therefore would not be represented in the EVS because the EVS does not currently include indels, nor does it list the location of intronic variants relative to the coding sequence (Exome Variant Server). Out of the 208 remaining variants, seven were present in the EVS (2.4% of the total variants). Four of these were present only in European Americans, two in African Americans, and one was present in both. Five of these seven were familial variants which were present in both affected and unaffected family members from the original referenced report by the ILAE, one variant was of unknown origin, and one was *de novo*. In comparison, 12% of the total ILAE annotated variants were familial but did not segregate with the disease, 16% were familial and segregated with the disease, 24% were *de novo*, and the remainder was of unknown origin. The frequency at which these seven variants were present in the EVS ranged from 7.6×10^{-5} to 0.008. Only a single variant, R221H in *EFHC1*, of the 290 present in the ILAE was present in the 1000 Genomes Project database, at a frequency of 0.018. The Polyphen scores for the seven variants ranged from benign to probably damaging (Table 2). Table 3 lists the frequency at which the various Polyphen scores appear for variants listed in the EVS for each gene. Additionally, we examined the EVS and the 1000 Genomes Project database for any nonsense variants in the 19 genes identified in the ILAE reports. In the EVS, we found five such variants among three genes, each at a low frequency (Table 4). No nonsense variants were found in the 1000 Genomes Project database.

Discussion

The variants examined in this study are found in the genes recommended for screening by the International League Against Epilepsy. Genetic testing for epilepsy can help a clinician make a diagnosis, eliminate the need for invasive diagnostic tests, inform treatment, and help families make reproductive decisions. However, genetic testing carries the risk of stigma, and it can affect employment and insurance opportunities. Therefore, it is vital that the role of a potentially pathogenic variant in causing an illness be verified before it is included in a clinical test so the correct clinical decisions can be made. Only seven (< 3%) of the variants delineated by the ILAE as epilepsy-associated were found in the 1000 Genomes Project database or the EVS. Five out of these seven variants were present in unaffected family members, so these mutations may have incomplete penetrance, which could explain their presence in the databases (Wallace, 2002). In the case of highly-penetrant epilepsy-associated variants, predictions based on small control sample sizes on whether variants are causal for epilepsy were reaffirmed when tested against a larger population.

Our analysis also revealed many potentially pathogenic variants in epilepsy genes in samples from the databases. Several of the variants that were present in the databases had *in vitro* evidence to support their pathogenic role. The C121W missense mutation in *SCN1B* resulted in a lower inactivation rate of sodium channels (Wallace et al., 1998), and all three of the variants identified in *EFHC1* decreased rates of apoptosis *in vitro* (Suzuki et al., 2004). In *SCN1A*, the T297I variant occurred in the pore-forming region of this protein, but affected a poorly conserved residue, (Nabbout et al., 2003), the E1957G variant occurred in the C terminus of the sodium channel, a region involved in association of the sodium channel with other proteins and its fast inactivation (Wallace et al., 2003), and the R1596C variant occurred in a highly conserved region (Harkin et al., 2007). Five nonsense variants were identified among the ILAE recommended epilepsy-associated genes in the EVS. These types of variants are likely to have a negative impact on the function of the gene, and harmful mutations are probably poorly tolerated among this set of genes. Finding functionally deleterious variants affecting genes known to play a role in monogenic epilepsies in the EVS database is consistent with recent reports demonstrating the existence of such variants in both neurologically normal controls (Klassen et al., 2011) and unaffected carriers (Wallace et al., 2002). The presence of functionally deleterious variants in the large exome databases has several explanations, including: 1) the variants may not be fully penetrant, but may play a modifier role in epilepsy as part of a complex genetic disease, 2) the patients in these databases were not necessarily free of epilepsy or other illnesses. Little phenotype information is available for the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010), and while the Exome Variant Server has associated phenotype information, it is not available for individual samples, so it is not possible to check if a rare variant is associated with a phenotype (Exome Variant Server). Moreover, the Exome Variant Server was created with the intention of identifying genes associated with heart, lung, and blood disorders. Since epilepsy is not one of the disorders investigated in these studies, it is possible that the subjects may have an undiagnosed or unreported seizure condition. Another issue with using these databases as control groups is that the filtering strategy used to create the 1000 Genomes Project was designed to include variants that have frequencies of at least 1%, so it may have excluded some pathogenic singletons (variants found in only one case).

None of the variants that were discovered in studies that segregated within larger families (defined as more affected individuals than just one parent and child) were present in the EVS, suggesting that existing criteria used to prove that highly penetrant variants are causal for epilepsy are robust. The present study confirms that the variants recently identified as

likely to play a role in epilepsy are largely absent from the general population, and demonstrates that the Exome Variant Server and the 1000 Genomes Project browser may be used as control groups to verify if a putative highly penetrant epilepsy-causal variant is present in the general population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010), NINDS NIH 1R01 NS064159-01A1 (to AGB). We thank Dr. Jeff Murray for carefully reviewing the manuscript.

References

- The 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
- Exome Variant Server. NHLBI Exome Sequencing Project (ESP); Seattle, WA: Retrieved July, 2012, from <http://evs.gs.washington.edu/EVS/>
- Harkin LA, McMahon JM, Iona X, Dibbens L, Pelekanos JT, Zuberi SM, et al. The spectrum of SCN1A-related infantile epileptic encephalopathies. *Brain*. 2007; 130(3):843–852. [PubMed: 17347258]
- Klassen T, Davis C, Goldman A, Burgess D, Chen T, Wheeler D, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell*. 2011; 145:1036–1048. [PubMed: 21703448]
- Mulley JC, Scheffer IE, Petrou S, Dibbens LM, Berkovic SF, Harkin LA. SCN1A mutations and epilepsy. *Hum. Mutat.* 2005; 25:535–542. [PubMed: 15880351]
- Nabbout R, Gennaro E, Dalla Bernardina B, Dulac O, Madia F, Bertini E, et al. Spectrum of SCN1A mutations in severe myoclonic epilepsies of infancy. *Neurology*. 2003; 60:1961–1967. [PubMed: 12821740]
- Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, et al. An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science*. 2012; 337(6090):100–104. [PubMed: 22604722]
- Ottman R, Hirose S, Jain S, Lerche H, Lopes-Cendes I, Noebels JL, et al. Genetic testing in the epilepsies—Report of the ILAE Genetics Commission. *Epilepsia*. 2010; 51:655–670. [PubMed: 20100225]
- Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, et al. Mutations in *EFHC1* cause juvenile myoclonic epilepsy. *Nat. Genet.* 2004; 36(8):842–849. [PubMed: 15258581]
- Tennessen JA, Bigam AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012; 337(6090):64–69. [PubMed: 22604720]
- Wallace RH, Wang DW, Singh R, Scheffer IE, George AL Jr, Phillips HA, et al. Febrile seizures and generalized epilepsy associated with a mutation in the Na⁺-channel β 1 subunit gene *SCN1B*. *Nat. Genet.* 1998; 19:366–370. [PubMed: 9697698]
- Wallace RH, Hodgson BL, Grinton BE, Gardiner RM, Robinson R, Rodriguez-Casero V, et al. Sodium channel α 1-subunit mutations in severe myoclonic epilepsy of infancy and infantile spasms. *Neurology*. 2003; 61:765–769. [PubMed: 14504318]

Table 1

Genes containing epilepsy-associated variants.

Gene	Product
<i>KCNQ2</i>	K _v 7.2 (K ⁺ channel)
<i>KCNQ3</i>	K _v 7.3 (K ⁺ channel)
<i>SCN2A</i>	Na _v 1.2 (Na ⁺ channel)
<i>STXBP1</i>	Syntaxin binding protein 1
<i>SCN1A</i>	Na _v 1.1 (Na ⁺ channel)
<i>SCN1B</i>	β ₁ subunit (Na ⁺ channel)
<i>GABRG2</i>	γ ₂ subunit (GABA _A receptor)
<i>SLC2A1</i>	GLUT1 (glucose transporter type 1)
<i>GABRA1</i>	α ₁ subunit (GABA _A receptor)
<i>EFHC1</i>	EF hand motif protein
<i>CHRNA4</i>	α ₄ subunit (nACh receptor)
<i>CHRNA2</i>	β ₂ subunit (nACh receptor)
<i>CHRNA2</i>	α ₂ subunit (nACh receptor)
<i>LGI1</i>	Leucine-rich repeat protein
<i>KCNMA1</i>	K _{Ca} 1.1 (K ⁺ channel)
<i>SLC2A1</i>	GLUT1 (glucose transporter type 1)
<i>CACNA1A</i>	Ca _v 2.1 (Ca ²⁺ channel)
<i>KCNA1</i>	K _v 1.1 (K ⁺ channel)
<i>ATP1A2</i>	Sodium-potassium ATPase

Table 2

Characteristics of epilepsy-associated variants that were found in the EVS database.

Gene	SCN1A	SCN1A	SCN1A	SCN1B	EFHCI	EFHCI	EFHCI
Mutation	T297I	E1957G	R1596C	C121W	F229L	P77T	R221H
De Novo?	familial	unknown	<i>De novo</i>	familial	familial	familial	familial
Segregates with disease in family?	no			no	no	no	no
Frequency In EVS	1/13001	5/13001	1/13005	1/13005	43/12963	107/12899	105/12901
European American	1/8595	5/8595	1/8599	1/8599	41/8559	0/8600	0/8600
African American	0/4406	0/4406	0/4406	0/4406	2/4404	107/4299	105/4301
PhastCons	0.068	1	1	1	1	0.103	0.905
GERP	5.41	5.67	5.9	3.17	6.01	3.13	2.2
Grantham Score	89	98	180	215	22	38	29
PolyPhen	probably-damaging	benign	probably-damaging	probably-damaging	possibly-damaging	benign	benign

The PhastCons score reflects the probability that a base is conserved between 17 vertebrate species. The Genomic Evolutionary Rate Profiling Score (GERP) is another measure of conservation ranging from -12.3 to 6.17, with 6.17 being the most conserved. The Grantham score ranks codon replacement by increasing chemical dissimilarity. The PolyPhen prediction uses the Polymorphism Phenotyping program to predict the possible impact of an amino acid substitution on a protein.

Table 3

Frequency of each Polyphen score among variants listed in the EVS for each gene.

Gene	Polyphen Score			
	benign	possibly-damaging	probably-damaging	unknown
<i>KCNQ2</i>	15	4	5	123
<i>KCNQ3</i>	23	8	17	122
<i>SCN2A</i>	27	5	14	167
<i>STXBPI</i>	7	2	6	115
<i>SCN1A</i>	30	21	28	151
<i>SCN1B</i>	10	4	7	42
<i>GABRG2</i>	5	1	2	68
<i>SLC2A1</i>	11	3	7	82
<i>GABRA1</i>	1	0	3	35
<i>EFHC1</i>	19	13	26	63
<i>CHRNA4</i>	20	13	23	106
<i>CHRN2</i>	7	3	12	46
<i>CHRNA2</i>	14	8	20	63
<i>LGII</i>	13	2	4	50
<i>KCNMA1</i>	15	7	8	190
<i>SLC2A1</i>	11	3	7	82
<i>CACNA1A</i>	41	15	21	285
<i>KCNA1</i>	14	2	1	30
<i>ATPIA2</i>	16	5	7	181

Gene	Product
<i>KCNQ2</i>	K _v 7.2 (K ⁺ channel)
<i>KCNQ3</i>	K _v 7.3 (K ⁺ channel)
<i>SCN2A</i>	Na _v 1.2 (Na ⁺ channel)
<i>STXBPI</i>	Syntaxin binding protein 1
<i>SCN1A</i>	Na _v 1.1 (Na ⁺ channel)
<i>SCN1B</i>	β ₁ subunit (Na ⁺ channel)
<i>GABRG2</i>	γ ₂ subunit (GABA _A receptor)
<i>SLC2A1</i>	GLUT1 (glucose transporter type 1)
<i>GABRA1</i>	α ₁ subunit (GABA _A receptor)
<i>EFHC1</i>	EF hand motif protein
<i>CHRNA4</i>	α ₄ subunit (nACh receptor)
<i>CHRN2</i>	β ₂ subunit (nACh receptor)
<i>CHRNA2</i>	α ₂ subunit (nACh receptor)
<i>LGII</i>	Leucine-rich repeat protein
<i>KCNMA1</i>	K _{Ca} 1.1 (K ⁺ channel)
<i>SLC2A1</i>	GLUT1 (glucose transporter type 1)

Gene	Product
<i>CACNA1A</i>	Ca _v 2.1 (Ca ²⁺ channel)
<i>KCNA1</i>	K _v 1.1 (K ⁺ channel)
<i>ATPIA2</i>	Sodium-potassium ATPase

Table 4

Frequency of stop-gained variants in genes linked to epilepsy.

Gene	Mutation	Frequency
<i>EFHC1</i>	Agr216X	1/13005
	Arg352X	2/13004
	Arg538X	1/13005
<i>CHRNA4</i>	Gln172X	2/13004
<i>CHRNA2</i>	Tyr331X	1/13005

These variants have not been previously identified as pathogenic